Dipole trap model for the metallic state in gated silicon-inversion layers

T. Hörmann

Institute for Semiconductor Physics, Johannes Kepler University, 4040 Linz, Austria; and
Christian Doppler Labor for Surface Optics, Johannes Kepler University, 4040 Linz, Austria

G. Brunthaler[*]

Institute for Semiconductor Physics, Johannes Kepler University, 4040 Linz, Austria

In order to investigate the metallic state in high-mobility Si-MOS structures, we have further developed and precised the dipole trap model which was originally proposed by B. L. Altshuler and D. L. Maslov [Phys. Rev. Lett. 82, 145 (1999)]. Our additional numerical treatment enables us to drop several approximations and to introduce a limited spatial depth of the trap states inside the oxide as well as to include a distribution of trap energies. It turns out that a pronounced metallic state can be caused by such trap states at appropriate energies whose behavior is in good agreement with experimental observations.

I. INTRODUCTION

The discovery of the metal-insulator transition (MIT) in two-dimensional (2D) electron systems in 1994[1,2] has attracted large attention, as it was in apparent contradiction to the scaling theory of localization[3,4] which states that in the limit of zero temperature, a metallic state should exist only in three dimensional systems, whereas in two dimensions disorder should always be strong enough to induce an insulating state. The MIT in high-mobility n-type silicon inversion layers shows a strong decrease of resistivity ρ towards low temperature T for high electron densities, manifesting the metallic region, whereas a strong exponential increase of ρ demonstrated the insulting regime for low densities. A similar but weaker behavior was observed in many other semiconductor systems at low densities and low temperatures (e.g. p-GaAs,[5] n-GaAs,[6] SiGe,[7] AlAs[8])

Several models were suggested in order to explain the unexpected finding of metallic behavior in 2D. The most important ones are i) temperature-dependent screening,[9–14] ii) quantum corrections in the diffusive regime,[15–18] and iii) quantum corrections in the ballistic regime.[19,20] Numerous argumentations for the different models are given in literature,[21–25] but a clear decision for one of them could not been drawn yet.

As an alternative, Altshuler and Maslov (AM) introduced the dipol scattering scenario for Si-MOS structures in which charged trap states in the oxide layer form dipols together with the image charge of the screening 2D electrons.[26] The interplay between the gate voltage dependent energetic position of the trap states and the height of the chemical potential may lead as well to a metal-insulator transition in that system. It should not be assumed that the dipol scattering effect is active alone, as the temperature dependence of screening and quantum corrections will surely contribute at low temperatures, but the charging of trap states might be the generator of the particularly large effect in Si-MOS structures. It is known that the misfit at the silicon/silicon-oxide interface produces charged defect states in the thermally grown oxide layer.[27–29] Arguments on the importance of trap states in Si-MOS structures were also given by Klapwijk and Das Sarma.[30]

AM could show within their analytical calculations that a trap level at energy E_T which is either filled or empty, depending on its position relative to the Fermi energy E_F, can lead to a critical behavior in electron scattering if E_T and E_F are degenerate. This dipole trap model is able to explain the main properties of the metal-insulator transition in gated Si-MOS structures.

For the analytical calculations AM made a number of assumptions. These are:

a1) the trap states possess a δ-like distribution in energy (i. e. have all the same energy),

a2) the spatial density distribution in the oxide is homogeneous,

a3) the states occupied with electrons behave neutral and cause no scattering of 2D electrons whereas the unoccupied states are positively charged and lead to scattering (AM work in the hole trap picture, but we describe occupation in terms of electrons),

a4) a charged trap state is screened by the 2D electrons so that the resulting electrostatic potential can be described by the trap charge and an apparent mirror charge with opposite sign on the other side of the interface,

a5) the scattering efficiency of the 2D electrons is described by a dipole field of the trap charge and its mirror charge,

a6) a parabolic saddle-point approximation for the effective potential between the Si/SiO2 interface and the metallic gate was used in order to perform analytical calculations,

a7) the energy of the trap state E_T is fixed relative to the quantization energy E_0 of the 2D ground state inside the inversion potential, and

a8) the chemical potential μ in the 2D layer has (A) either the same temperature dependence as in the bulk

substrate or (B) as in a 2D electron system with constant electron density.

In this work, we precise and develop the AM trap model further in order to better understand the influence of charged traps on the metallic state in Si-MOS structures. We present detailed numerical calculations of the temperature and density dependent resistivity ρ due to electronic scattering in the dipole trap model. Due to the numerical treatment we were able to drop the approximations a6), a7), and a8) of the analytic AM model. In addition, we have further extended our calculations for the more realistic case with energetic broadening and spatial distribution profile of the defect states, i. e. dropping also approximations a1) and a2).

As a result of our calculations, we find good agreement with the calculations of AM. There are mainly deviations in the overall behavior of the resistivity at low electron densities and at high temperatures. In order to understand the approximations of AM, we have also recalculated the analytical model and were able to formulate it in a simplified way. Several mathematical terms are rearranged so that the scattering efficiency is expressed in the same form as in the usual Drude formulation. In addition, their result contains two integrals, which we could replace by Fermi-Dirac integrals. Thus the known approximations for the Fermi-Dirac integrals lead to simple equations for the resistivity ρ at low temperatures.

The paper is organized as follows. In Sec. II the analytical formulation of the dipole trap model is given in detail and we show that within the saddle-point approximation the result can be written in terms of Fermi-Dirac integrals. Section III treats the chemical potential and Sec. IV the analytic approximations for low temperatures. From the numerical integration, first results are given in Sec. V whereas in Sec. VI the calculations are extended for the case that the conduction band is the reference energy for the trap energy and not the electronic ground state E_0. In Sec. VII the model and the calculation are extended for a spatial density distribution of the trap states and in Sec. VIII energetic broadening of trap states is taken into account. Conclusions are drawn in Sec. IX. In two appendices the behavior of the chemical potential and the ground state energy of the inversion layer are described in detail. Please note that we use SI units throughout this work.

II. TRAPMODEL

In the AM model it is assumed that a large number of hole trap states exists in the oxide at a certain trap energy E_T. If the trap energy lies above the chemical potential μ, the trap is empty (in the electron picture, or has captured a hole in the equivalent description) and is positively charged, whereas if E_T lies below μ it is filled with an electron and thus is neutral, see Fig. 1. Please note that we use in this work the terminology $\mu(T)$ for the chemical potential, the Fermi energy E_F
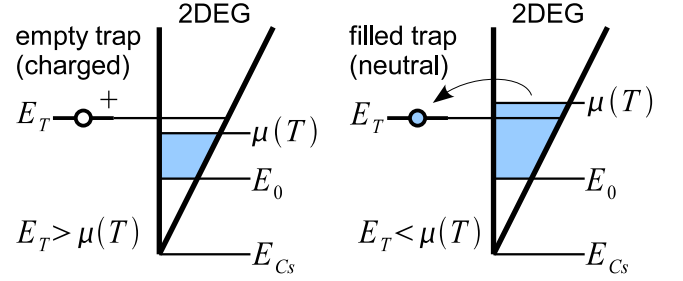

Figure 1. Schematic representation of the trap states together with the 2D electron system in the Si inversion layer. For $E_T > \mu(T)$ the trap state is positively charged and scatters electrons in the 2D layer whereas for $E_T < \mu(T)$ the trap is neutral.

denotes $\mu(T = 0)$.

A potential gradient due to an applied gate voltage V_g causes a decrease of the trap energy $E_T = E_{Ts} - eV_{\text{ins}}Z/D$, with the unscreened trap energy E_{Ts} at the oxide semiconductor (OS) interface, the voltage drop across the oxide (insulator) V_{ins}, the distance from the OS interface Z, and the thickness of the oxide D. In their corresponding equation AM use the symbol V_g instead of V_{ins}. . But the total gate voltage V_g is equal to the voltage drop across the oxide V_{ins} plus the voltage drop across the depletion layer V_{depl}. Later in their paper AM write that the threshold voltage is incorporated into V_g, nevertheless they use practically an equation equivalent to

$$V_{\text{ins}} = e n_s D / \epsilon_{\text{ins}} \epsilon_0, \tag{1}$$

where ϵ_{ins} is the dielectric constant of the oxide and ϵ_0 is the electric field constant. But if the threshold voltage is incorporated into V_g, the latter cannot be used to calculate the slope of the electrical potential within the oxide, as a part of V_g falls off between the OS interface and the bulk layer. As the charges within the depletion layer (2D charge density $-e n_{\text{depl}}$) also contribute to the gradient of the potential, we use the equation

$$V_{\text{ins}} = e \left(n_s + n_{\text{depl}} \right) D / \epsilon_{\text{ins}} \epsilon_0 \tag{2}$$

together with

$$V_g = V_{\text{depl}} + V_{\text{ins}}. \tag{3}$$

According to AM another term has to be added to the trap energy E_T which describes the interaction with the two dimensional electron gas (2DEG) because of the image force. In Ref. 26 this term is given as $-e^2/8\pi\epsilon_{\text{ins}}\epsilon_0 Z$ (translated to the SI unit system). We think there should be a factor 16 in the denominator, see for instance Ref. 27 or Ref. 28. This factor does not change the results of the trap model qualitatively therefore we write $-e^2/C\pi\epsilon_{\text{ins}}\epsilon_0 Z$.

AM define an energy

$$\varepsilon_D = \frac{e^2}{C\pi\epsilon_{\text{ins}}\epsilon_0 D} \tag{4}$$
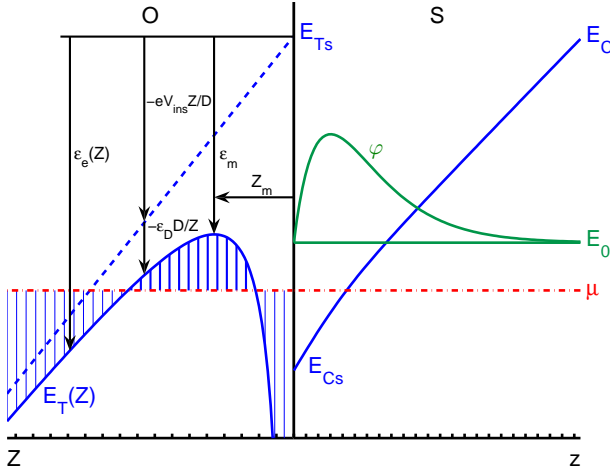
Figure 2. Oxide semiconductor interface. For simplicity we use different coordinate systems for the oxide (Z) and the semiconductor side (z) so that both, Z and z, are positive on their sides. Trap energy E_T which reaches its maximum at $Z = Z_m$, unscreened trap energy E_{Ts} at the OS interface, electrostatic energy ε_e, its components $-eV_{\mathrm{ins}}Z/D$ and $-\varepsilon_D D/Z$, and its maximum ε_m, chemical potential μ, ground state energy of the inversion layer E_0 and the corresponding wave function φ, conduction band edge E_C and its value at the interface E_{Cs}.

with $C = 8$ (AM) or $C = 16$ (our assumption). (We use in this work the notation of capital E for absolute energies and Greek ε for energy differences.) Finally the trap energy can be written as

$$E_T (Z) = E_{Ts} + \varepsilon_e (Z), \qquad (5)$$

$$\varepsilon_e (Z) = -eV_{\mathrm{ins}}\frac{Z}{D} - \varepsilon_D \frac{D}{Z}, \qquad (6)$$

where the subscript e in ε_e stands for 'electrostatic'. The shape of $E_T (Z)$ is shown in Fig. 2 together with the relevant energy notation.

Without the presence of a magnetic field the probability of a trap to be charged is given by

$$p_+ (Z) = \frac{1}{\frac{1}{2} \exp\left(-\frac{E_T(Z)-\mu}{k_B T}\right) + 1}. \qquad (7)$$

This formula is similar to the Fermi distribution function, differences are the minus sign in the exponent and the prefactor $1/2$. A trap is charged when it is not occupied by an electron, thus the minus sign. If the trap is charged there are two possibilities for the spin orientation but only one if it is not charged, leading to the factor $1/2$. As mentioned before p_+ is determined by $E_T - \mu$ (the vertical lines in Fig. 2) and the temperature.

AM assume that a positive charged trap is screened by the electrons in the 2DEG and the trap forms together with that image charge a dipole. For the transport scat-

tering cross section σ_t of this dipole they found classically

$$\sigma_t (\varepsilon, Z) = 2.74 \left(\frac{e^2 Z^2}{8\pi\epsilon^*\epsilon_0\varepsilon}\right)^{1/3} \equiv c_\sigma \varepsilon^{-1/3} Z^{2/3}, \qquad (8)$$

$$\rightarrow \quad c_\sigma = 2.74 \left(\frac{e^2}{8\pi\epsilon^*\epsilon_0}\right)^{1/3}, \qquad (9)$$

where Z is the distance between the trap and the oxide semiconductor interface, ε is the energy of the scattered electron relative to the ground state energy of the inversion layer E_0:

$$\varepsilon = E - E_0, \qquad (10)$$

and ϵ^* is an effective dielectric constant

$$\epsilon^* = \frac{\epsilon_{\mathrm{ins}} + \epsilon_{\mathrm{sc}}}{2}, \qquad (11)$$

where ϵ_{sc} is the dielectric constant of the semiconductor. We have recalculated Eq. (8) and got the same result as AM.

The Drude formula together with the Boltzmann equation in relaxation time approximation yields the resistivity $d\rho(Z)$ caused by the charged traps within the layer $[Z, Z + dZ]$. By integrating this contributions over the whole oxide $[0, D]$ one gets

$$\rho = \frac{\sqrt{2m_c}c_\sigma\bar{\varepsilon}^{1/6}\int_0^D N_T^+ Z^{2/3}dZ}{n_s e^2}, \qquad (12)$$

$$\bar{\varepsilon} = \varepsilon_{F0} \left[\int_0^\infty \frac{1}{4k_B T}\left(\frac{\varepsilon}{\varepsilon_{F0}}\right)^{5/6} \cosh^{-2}\left(\frac{\varepsilon - \mu_{E_0}}{2k_B T}\right) d\varepsilon\right]^{-6}, \qquad (13)$$

where m_c is the conductivity mass of the free electrons within the inversion layer and $\bar{\varepsilon}$ is an effective electron energy as used by AM. In the corresponding AM equation the argument of the cosh function contains the Fermi energy, but should be replaced by the chemical potential.[31] Furthermore $\varepsilon_{F0} = E_F - E_0$ is the Fermi energy, $\mu_{E_0} = \mu - E_0$ is the chemical potential, each relative to the ground state energy of the inversion layer, and N_T^+ is the 3D density of charged traps. With the 3D trap density $N_T (Z)$ it is given by

$$N_T^+ (Z) = N_T (Z) p_+ (Z). \qquad (14)$$

Now we like to present Eq. (12) in a form similar to the Drude formula

$$\rho = \frac{m_c}{n_s e^2 \langle\tau\rangle}, \qquad (15)$$

and further find a term for the scattering rate $1/\langle\tau\rangle$ in the usual form that scattering rate is equal to scattering cross section times density of scattering centers times velocity of scattered particles, which leads to one of the basic equations used throughout this work

$$\rho = \frac{m_c\sigma_t\left(\bar{\varepsilon}, \bar{Z}\right)n_T^+ v\left(\bar{\varepsilon}\right)}{n_s e^2}. \qquad (16)$$

Here $\sigma_t\left(\bar{\varepsilon}, \bar{Z}\right)$ is the transport scattering cross section from Eq. (8) for the effective electron energy $\bar{\varepsilon}$ and an effective distance $\bar{Z}$ between the traps and the OS interface, $n_T^+ = \int_0^D N_T^+\left(Z\right) dZ$ is the 2D density of charged traps, and $v\left(\bar{\varepsilon}\right)$ is the electron velocity which corresponds with $\bar{\varepsilon} = m_c v^2/2$. Comparing Eq. (12) and (16) we find

$$\bar{Z}^{2/3} = \frac{\int_0^D N_T^+\left(Z\right) Z^{2/3} dZ}{\int_0^D N_T^+\left(Z\right) dZ} = \frac{\int_0^D N_T^+\left(Z\right) Z^{2/3} dZ}{n_T^+} \quad (17)$$

for this factor in $\sigma_t\left(\bar{\varepsilon}, \bar{Z}\right) = c_\sigma \bar{\varepsilon}^{-1/3} \bar{Z}^{2/3}$.

In order to calculate the resistivity ρ the knowledge of n_T^+ is not necessary, n_T^+ cancels out with the denominator of $\bar{Z}^{2/3}$ within $\sigma_t\left(\bar{\varepsilon}, \bar{Z}\right)$, AM do not use it. As mentioned in the introduction we are interested in the metal-insulator transition depending on the electron density n_s, i.e. we want to know the temperature behavior of ρ as a function of n_s. In this context n_T^+ is very useful in order to understand on the basis of Eq. (16) that it contributes the main variations to the resistivity $\rho\left(n_s, T\right)$ whereas $\sigma_t\left(\bar{\varepsilon}, \bar{Z}\right)$ and $v\left(\bar{\varepsilon}\right)$ show only weak dependence on n_s and T. The benefit of Eq. (16) against (12) is, that the physical meaning of the terms is immediately clear.

The integral in the numerator and that in the denominator of $\bar{Z}^{2/3}$ can be treated in quite the same way, so we define

$$\Omega_j \equiv \int_0^D N_T^+\left(Z\right) Z^j dZ =$$

$$= N_T \int_0^D \frac{Z^j dZ}{\frac{1}{2}\exp\left(-\frac{E_T(Z)-\mu}{k_B T}\right)+1}. \quad (18)$$

In the last step we followed AM and assumed that the trap density is constant within the oxide, respectively in the region where $p_+\left(Z\right)$ does not vanish. Now we can write

$$\sigma_t\left(\bar{\varepsilon}, \bar{Z}\right) = c_\sigma \bar{\varepsilon}^{-1/3} \bar{Z}^{2/3} = c_\sigma \bar{\varepsilon}^{-1/3}\frac{\Omega_{2/3}}{\Omega_0}, \quad (19)$$

$$n_T^+ = \Omega_0, \quad (20)$$

$$\rho = \frac{\sqrt{2m_c} c_\sigma \bar{\varepsilon}^{1/6}\Omega_{2/3}}{n_s e^2}. \quad (21)$$

To be able to calculate the integral which corresponds with $\Omega_{2/3}$ AM expanded the electrostatic energy $\varepsilon_T\left(Z\right)$ into a Taylor series about the point Z_m where it reaches its maximum ε_m. This procedure is called saddle-point approximation.

$$Z_m = D\sqrt{\frac{\varepsilon_D}{eV_{\text{ins}}}}, \quad (22)$$

$$\varepsilon_m = -2\sqrt{eV_{\text{ins}}\varepsilon_D}, \quad (23)$$

$$\varepsilon_e\left(Z\right) \simeq \varepsilon_m - \varepsilon_D \frac{D}{Z_m^3}\left(Z - Z_m\right)^2, \quad (24)$$
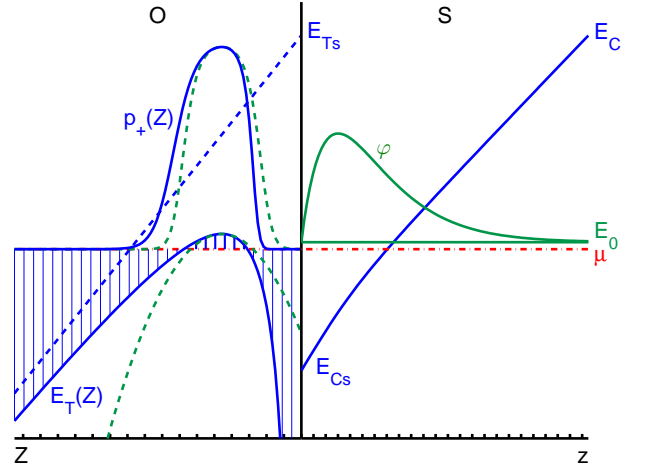

Figure 3. Oxide-semiconductor interface. Trap energy $E_T\left(Z\right)$ (full line), its Taylor approximation (dashed line), and the resulting probabilities $p_+\left(Z\right)$ in arbitrary units with μ as zero point.

see Fig. 2 and 3. Now (18) can be written as

$$\Omega_j \simeq N_T \int_0^D \frac{Z^j dZ}{\frac{1}{2}\exp\left(\frac{\mu_{E_0}-\varepsilon_{Ts0}-\varepsilon_m+\varepsilon_D \frac{D}{Z_m^3}(Z-Z_m)^2}{k_B T}\right)+1}, \quad (25)$$

$$\varepsilon_{Ts0} = E_{Ts} - E_0 = \text{const.} \quad (26)$$

AM assume that the energy E_{Ts} relative to the ground state energy E_0 is constant, but we believe that rather the conduction band edge at the interface E_{Cs} has to be used as reference energy, i.e. $\varepsilon_{TsCs} = E_{Ts} - E_{Cs} = \text{const.}$ This issue will be further treated in section VI.

The integrand is a peak around Z_m which drops off exponentially on both sides. In order to come to the same result as AM, we further apply the following approximations: (i) Because of the exponential decrease one can integrate from $-\infty$ to ∞. (ii) The integrand is dominated by the denominator, so $Z^j \simeq Z_m^j$ can be set in the numerator. Now the integrand is symmetric around Z_m and with help of the substitution $\mathcal{Z} = \frac{\varepsilon_D D}{Z_m^3 k_B T}\left(Z - Z_m\right)^2$ one gets

$$\Omega_j = N_T Z_m^{j+3/2}\sqrt{\frac{k_B T}{\varepsilon_D D}}$$

$$\times \int_0^\infty \frac{\mathcal{Z}^{-1/2} d\mathcal{Z}}{\exp\left(\frac{\mu_{E_0}-\varepsilon_{Ts0}-\varepsilon_m}{k_B T}-\ln 2+\mathcal{Z}\right)+1}. \quad (27)$$

This corresponds to the integral in equation (9c) in Ref. 26 ($\mathcal{Z}$ corresponds to x^2). We brought it into the form above as it corresponds now to a Fermi-Dirac integral[32]

$$\mathcal{F}_k\left(\eta\right) = \frac{1}{\Gamma\left(k+1\right)}\int_0^\infty \frac{\mathcal{Z}^k d\mathcal{Z}}{\exp\left(\mathcal{Z}-\eta\right)+1}, \quad (28)$$

where Γ is the Gamma function. A comparison yields

$$\Omega_j = N_T Z_m^{j+3/2} \sqrt{\frac{k_B T}{\varepsilon_D D}}$$
$$\times \underbrace{\sqrt{\pi}}_{\Gamma\left(\frac{1}{2}\right)} \mathcal{F}_{-1/2}\left(\ln 2 + \frac{\varepsilon_{Ts0} + \varepsilon_m - \mu_{E_0}}{k_B T}\right). \quad (29)$$

On the right hand side j appears only in the exponent of Z_m, so within the saddle-point approximation Eq. (17) simplifies to

$$\bar{Z}^{2/3} = \frac{\Omega_{2/3}}{\Omega_0} \simeq Z_m^{2/3}. \quad (30)$$

III. CHEMICAL POTENTIAL

AM described two scenarios for the temperature behavior of the chemical potential: (A) The chemical potential of the 2DEG and of the Si substrate coincide. (B) The 2DEG is disconnected form the substrate. For the case (A) they assumed that the temperature behavior in the 2DEG is the same as in the bulk. However, they did not take into account that the chemical potential in the 2DEG is measured against the ground state energy E_0 and in the bulk against the conduction or valence band edge, i.e. they assumed E_0 and the band bending to be fixed.

For (B) AM used an equation analogous to

$$\mu_{E_0} = k_B T \ln\left[\exp\left(\frac{\varepsilon_{F0}}{k_B T}\right) - 1\right]. \quad (31)$$

If only one subband of the inversion layer is occupied (quantum limit), the Fermi energy relative to its ground state energy is given by[28]

$$\varepsilon_{F0} = \frac{2\pi\hbar^2 n_s}{g_s g_{v2D} m_{d2D}}, \quad (32)$$

where g_s, g_{v2D}, and m_{d2D} are the spin degeneracy, the valley degeneracy (for the 2DEG), and the density-of-states mass (2D) respectively. The two equations above can be derived from

$$n_s = \underbrace{\frac{g_s g_{v2D} m_{d2D}}{2\pi\hbar^2}}_{\text{2D density of states}} \int_{E_0}^{\infty} \underbrace{\frac{1}{\exp\left(\frac{E-\mu}{k_B T}\right) + 1}}_{\text{Fermi$-$Dirac distribution}} dE. \quad (33)$$

Our assumptions are shown schematically in Fig. 4 (semiconductor side of the OS interface) and are as follows. In thermal equilibrium there is a single chemical potential μ throughout the structure. For a certain temperature T, μ is determined in the bulk by the (residual) doping density (giving $\mu_{Cb} = \mu - E_{Cb}$, for details see App. B). In the inversion layer the position of μ relative to E_0 (i.e. μ_{ε_0}) follows directly from the 2D density n_s.
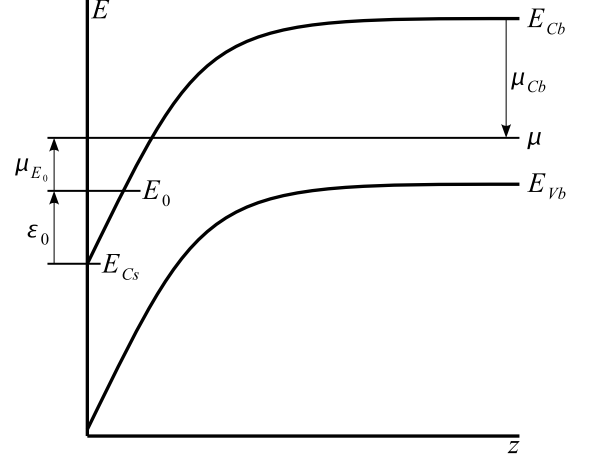

Figure 4. Band bending and notation on semiconductor side of the OS interface for thermal equilibrium between 2D layer and bulk. The electronic ground state energy of the inversion layer E_0, the conduction band edge E_C, its values at the interface E_{Cs} and in the bulk E_{Cb}, the valence band edge E_V, and the chemical potential μ are shown schematically, also the ground state energy relative to the conduction band edge ε_0, the chemical potential relative to conduction band edge in the bulk μ_{Cb} and relative to the ground state energy μ_{E_0} are noted.

The band bending $E_{Cb} - E_{Cs}$ adjusts so that

$$E_{Cb} - E_{Cs} = \varepsilon_0 + \mu_{E_0} - \mu_{Cb}. \quad (34)$$

If the band bending is increased, the quantum well gets narrower and the ground state energy increases. Thus ε_0 itself is a function of the band bending, a self consistent calculation solves this problem (fix point iteration).

Another possibility is that the 2D electron system is at low temperatures decoupled from the bulk substrate and no common chemical potential exists. We do not treat this case in this work, but expect a similar behavior.

Fig. 5 shows the behavior of $\mu(T, n_s)$ which is crucial for the understanding of the behavior of $\rho(T, n_s)$. The chemical potential relative to the ground state energy of the inversion layer $\mu_{E_0}(T, n_s)$ increases for decreasing T and for increasing n_s.

The parameters, which were used here and for the following calculations are collected in Table I. We assumed a {001} plane for the silicon surface.

IV. APPROXIMATION FOR LOW TEMPERATURES

For given values T and n_s we calculate the resistivity ρ with help of the equations (21) and (29). Beside the explicit temperature dependence also the chemical potential μ_{E_0} and the effective electron energy $\bar{\varepsilon}$ are functions of T.
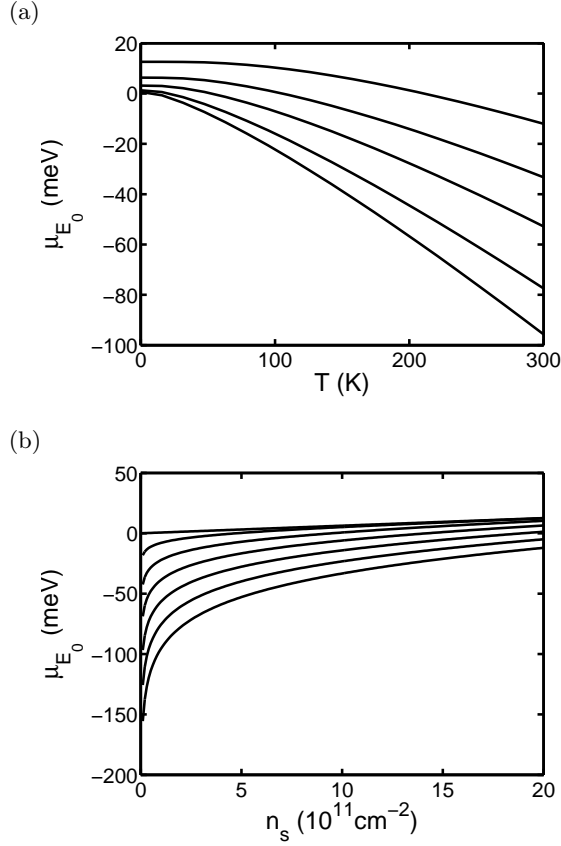
As a first step we replace the integral in (13) by a

(a)

(b)

Figure 5. Chemical potential μ relative to the ground state energy E_0. a) Curves with constant n_s, bottom up: $n_s = 1, 2, 5, 10, 20 \times 10^{11}\,\text{cm}^{-11}$. b) Curves for constant T, top down: $T = 0, 50, 100, 150, 200, 250, 300\,K$.

Table I. Values used for calculations.

$$g_s = 2$$
$$g_{\text{v2D}} = 2$$
$$g_{\text{v3D}} = 6$$
$$m_e = 9.1094 \times 10^{-31}\,\text{kg}$$
$$m_{\text{d2D}} = 0.1905\,m_e$$
$$m_z = 0.9163\,m_e$$
$$m_{\text{de3D}} = 0.322\,m_e$$
$$m_{\text{dh3D}} = 0.59\,m_e$$
$$\epsilon_{\text{sc}} = 11.9$$
$$\epsilon_{\text{ins}} = 3.9$$
$$\rightarrow \epsilon^* = 7.9$$
$$D = 200\,\text{nm}$$
$$C = 16$$
$$\rightarrow \varepsilon_D = 0.4615\,\text{meV}$$
$$N_A = 2 \times 10^{15}\,\text{cm}^{-3}$$
$$N_D = 0$$
$$\varepsilon_{\text{g0Si}} = 1.17\,\text{eV}$$
$$\alpha_{\text{Si}} = 4.73 \times 10^{-4}\,\text{eVK}^{-1}$$
$$\beta_{\text{Si}} = 636\,\text{K}$$

Fermi-Dirac integral. For the Fermi distribution function

$$f = \frac{1}{\exp\left(\frac{E-\mu}{k_B T}\right) + 1} = \frac{1}{\exp\left(\frac{\varepsilon - \mu_{E_0}}{k_B T}\right) + 1} \tag{35}$$

the following identity holds,

$$-\frac{\partial f(\varepsilon)}{\partial \varepsilon} = \frac{1}{4 k_B T} \cosh^{-2}\left(\frac{\varepsilon - \mu_{E_0}}{2 k_B T}\right). \tag{36}$$

So Eq. (13) can be written as

$$\bar{\varepsilon} = \varepsilon_{F0} \left[\int_0^\infty \left(\frac{\varepsilon}{\varepsilon_{F0}}\right)^{5/6} \left(-\frac{\partial f(\varepsilon)}{\partial \varepsilon}\right) d\varepsilon \right]^{-6}. \tag{37}$$

An integration by parts yields

$$\bar{\varepsilon} = \varepsilon_{F0}^6 \left[\frac{5}{6} \int_0^\infty \varepsilon^{-1/6} f(\varepsilon)\, d\varepsilon\right]^{-6},$$

$$\bar{\varepsilon} = \varepsilon_{F0} \left(\frac{\varepsilon_{F0}}{k_B T}\right)^5$$

$$\times \left[\frac{5}{6} \int_0^\infty \frac{\left(\frac{\varepsilon}{k_B T}\right)^{-1/6}}{\exp\left(\frac{\varepsilon - \mu_{E_0}}{k_B T}\right) + 1} d\left(\frac{\varepsilon}{k_B T}\right)\right]^{-6},$$

$$\bar{\varepsilon} = \varepsilon_{F0} \left(\frac{\varepsilon_{F0}}{k_B T}\right)^5 \left[\Gamma\left(\frac{11}{6}\right) \mathcal{F}_{-1/6}\left(\frac{\mu_{E_0}}{k_B T}\right)\right]^{-6}. \tag{38}$$

For low temperatures we can take advantage of the behavior of $\mu_{E_0}(T)$ and $\bar{\varepsilon}(T)$ for $T \to 0$. It can easily be shown that for constant n_s and therefore constant ε_{F0} all derivatives vanish at this point.

$$\left.\frac{d^n \mu_{E0}(T)}{dT^n}\right|_{T \to 0} = 0 \quad \text{for } n \geq 1, \tag{39}$$

$$\left.\frac{d^n \bar{\varepsilon}(T)}{dT^n}\right|_{T \to 0} = 0 \quad \text{for } n \geq 1. \tag{40}$$

So $\mu_{E_0}(T)$ and $\bar{\varepsilon}(T)$ are very flat functions at $T \to 0$, for $k_B T \ll \varepsilon_{F0}$ they can be approximated by

$$\mu_{E_0}(T) \simeq \varepsilon_{F0}, \tag{41}$$

$$\bar{\varepsilon}(T) \simeq \varepsilon_{F0}. \tag{42}$$

See appendix A for details.

The Fermi-Dirac integral $\mathcal{F}_j(\eta)$ can be approximated by[32]

$$\mathcal{F}_j(\eta) \simeq \frac{\eta^{j+1}}{\Gamma(j+2)} \quad \text{for } \eta \gg 0, \tag{43}$$

which is the first term of an asymptotic series derived with help of a Sommerfeld expansion,[33] and by

$$\mathcal{F}_j(\eta) \simeq \exp \eta \quad \text{for } \eta \ll 0. \tag{44}$$

For $T \to 0$ the argument of the Fermi-Dirac integral in Eq. (29) increases beyond any border. Which approximation for the Fermi-Dirac integral is applicable depends

on the sign of $\varepsilon_{Ts0} + \varepsilon_m - \varepsilon_{F0}$. (Here μ_{E_0} is replaced by ε_{F0} as by definition $\mu_{E_0} \to \varepsilon_{F0}$ for $T \to 0$.) This is conform with AM's definition of the transition point, $(\varepsilon_{Ts0} + \varepsilon_m - \varepsilon_{F0})/k_B T = 0$. Accordingly we define

$$\varepsilon_{mF} = \varepsilon_{Ts0} + \varepsilon_m - \varepsilon_{F0},$$
$$\varepsilon_{mF} > 0 \quad \to \text{insulating},$$
$$\varepsilon_{mF} < 0 \quad \to \text{metallic},$$
$$\varepsilon_{mF} = 0 \quad \to \text{transition point}$$

and a critical density

$$n_{sc} = n_s|_{\varepsilon_{mF}=0}. \tag{45}$$

Applying the appropriate approximations results in

$$\Omega_j \simeq \begin{cases} 2N_T Z_m^{j+3/2} \sqrt{\dfrac{k_B T \ln 2}{\varepsilon_D D} + \dfrac{\varepsilon_{mF}}{\varepsilon_D D}} & \text{for } \varepsilon_{mF} > 0 \\[2ex] 2N_T Z_m^{j+3/2} \sqrt{\dfrac{k_B T \pi}{\varepsilon_D D} \exp\left(\dfrac{\varepsilon_{mF}}{k_B T}\right)} & \text{for } \varepsilon_{mF} < 0 \\[2ex] \underbrace{\mathcal{F}_{-1/2}(\ln 2)}_{\simeq 0.891} N_T Z_m^{j+3/2} \sqrt{\dfrac{k_B T \pi}{\varepsilon_D D}} & \text{for } \varepsilon_{mF} = 0. \end{cases}$$
$$\tag{46}$$

Please note an interesting behavior. When setting $\varepsilon_{mF} = 0$ in the first two equations for $T > 0$ they converge neither into each other nor into the third one. This apparent discrepancy can be understood, as $|\varepsilon_{mF}|$ gets smaller and smaller the maximum temperature where the approximations for the Fermi-Dirac integral are just applicable also gets smaller and smaller and finally vanishes for $\varepsilon_{mF} = 0$. Indeed for $\varepsilon_{mF} = 0$ and $T \to 0$ the three cases yield the same result, i. e. $\Omega_j = 0$.

V. COMPARISON OF ANALYTIC AND NUMERICAL RESULTS

In order to get rid of the restrictions which came from the saddle-point approximation we have performed numerical calculations of the integrals Ω_j (Eq. (25)). In Fig. 6 we show the resistivity ρ depending on the temperature T and the 2D electron density in the inversion layer n_s calculated with the help of the saddle-point approximation (full lines), the approximation for low temperatures (dashed lines), and the numerical integration (markers). We chose $\varepsilon_{Ts0} \simeq 42.02\,\text{meV}$ in order to get $n_{sc} = 10^{11}\text{cm}^{-11}$, for the 3D trap density we assumed $N_T = 10^{18}\,\text{cm}^{-3}$. The value for N_T seems to be quite high, but only in a very narrow layer the traps will indeed be charged (where the trap states are above the chemical potential) and contribute to scattering. A further limitation of the available trap states into a narrow region besides the interface will follow later in this work.

The critical behavior of $\rho(T)$ versus n_s for the homogeneous trap density N_T can be seen most clearly on the double logarithmic Figs. 6a) and c). We added 6b) as it is in this form directly comparable with Fig. 1b) in the work of AM.[26]

We like to explain the shape of $\rho(T, n_s)$ starting from a value $n_s > n_{sc}$. The maximum of $\varepsilon_T(Z)$ lies below μ, therefore the $p_+(Z)$-peak is very small and narrow. When n_s is decreased μ drops off (see Fig. 5) and the resistivity ρ rises fast because $-(E_T - \mu)/k_B T$ is the exponent in the denominator of p_+. It rises the faster the smaller the temperature is. When μ reaches the maximum of the trap energy E_T the p_+-peak has a height of $2/3$, therefore further decreasing of n_s and μ cannot increase the height of the peak appreciable (on a logarithmic scale, which covers some orders of magnitude), only the width.

For increasing temperature T also the chemical potential μ measured against E_0 decreases which results in increasing resistivity ρ. Additionally the $p_+(Z)$-peak is broadened because the Fermi distribution function declines over several $k_B T$.

The saddle-point approximation works best for small temperatures T and densities $n_s > n_{sc}$, in this domain the $p_+(Z)$-peak is very narrow and therefore the Taylor approximation of the trap energy is accurate within the peak. For large T and/or $n_s < n_{sc}$ the deviations of the saddle-point approximation from the numerical calculations is visible in Fig. 6.

It should be noted that according to the calculations, the resistivity ρ drops to arbitrarily low values in the metallic regime. This is caused by the fact that electron scattering is taken into account only by the trap states at a single trap energy. If this trap energy is below the Fermi energy, with decreasing T the number of charged scattering centers goes to zero. Only if other scattering effects, like residual impurities, surface roughness, acceptor states in the depletion layer, etc. are included, the low-T resistivity would be limited. As will be seen later, an energetic broadening of the trap states will have a similar effect.

VI. CONDUCTION BAND AS REFERENCE ENERGY

If the bands in the semiconductor and in the oxide are bent due to an applied gate voltage all energies move up and down with the bands. The energetic position of the trap states E_{Ts} should thus be fixed relative to the conduction band edge and not to the ground state energy E_0 of the inversion layer as assumed by AM.

From Eq. (7) we see that $E_T - \mu = E_{Ts} + \varepsilon_e(Z) - \mu$ determines the probability p_+ of a trap to be charged. So if we measure E_{Ts} against E_{Cs} we also have to know $\mu_{Cs} = \mu - E_{Cs}$. We find

$$\mu_{Cs} = \mu - E_0 + E_0 - E_{Cs} = \mu_{E_0} + \varepsilon_0, \tag{47}$$

where $\varepsilon_0 = E_0 - E_{Cs}$ is the ground state energy relative to the conduction band edge at the OS interface.

The chemical potential μ_{E_0} follows from Eq. (31), but an accurate calculation of ε_0 is rather complex. For simplicity we follow the calculation method of Ando, Fowler
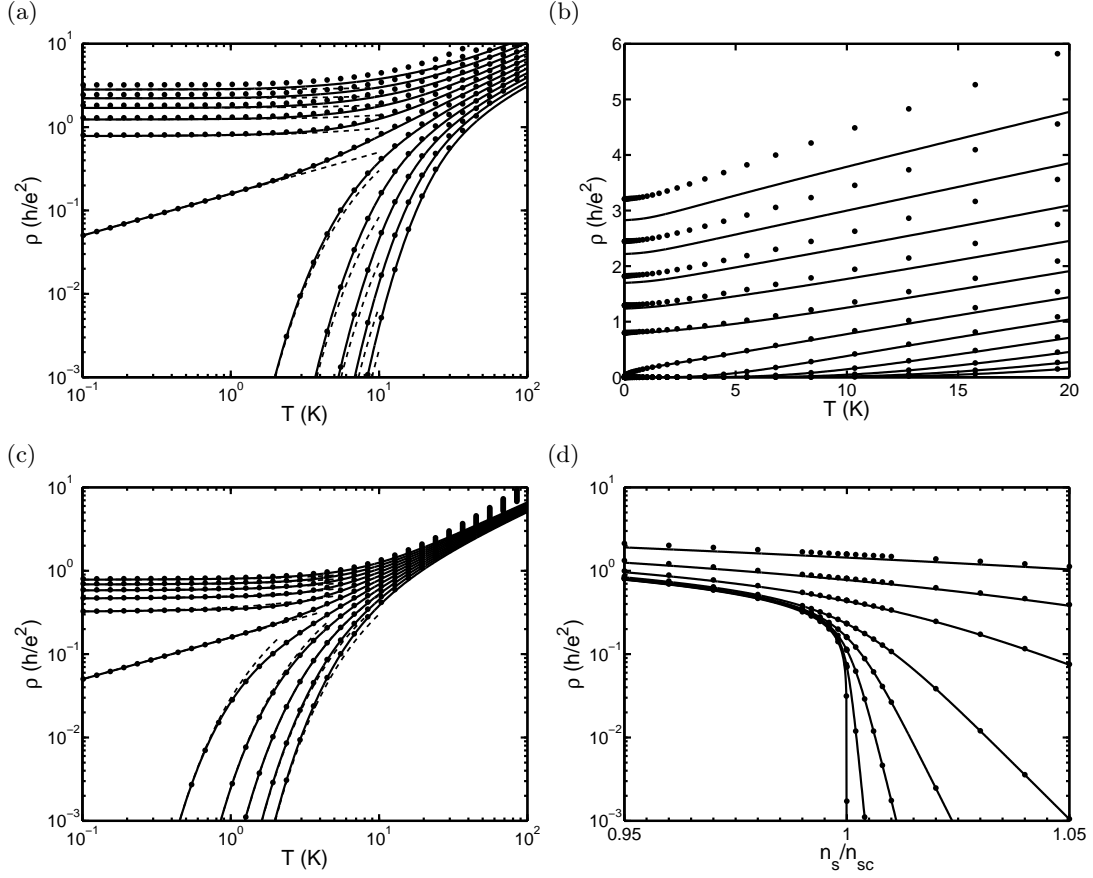
Figure 6. Behavior of the resistivity ρ depending on the temperature T and the 2D electron density in the inversion layer n_s. Full lines: saddle-point approximation, point symbols: numerical integration, dashed lines: low temperature approximations. On the curves in a) - c) n_s and in d) T is constant. Critical density $n_{sc} = 10^{11}\,\mathrm{cm}^{-2}$. a) logarithmic and b) linear display of $\rho(T)$ with parameters: $n_s = 0.75, \ldots, 1.25 \times n_{sc}$ in steps of $0.05 \times n_{sc}$ in top down order for the individual curves, c) logarithmic view of $\rho(T)$ with top down parameters: $n_s = 0.95, \ldots, 1.05 \times n_{sc}$ in steps of $0.01 \times n_{sc}$, d) $\rho(n_s)$ with parameters $T = 0,\ 0.2,\ 0.5,\ 1,\ 2,\ 5,\ 10,\ 20\,\mathrm{K}$ in bottom up order.

and Stern (AFS)[28] and neglect the exchange interaction and correlation effects and use the Ritz variational principle. (In the mentioned article also more sophisticated methods for the calculation of ε_0 are given.)

For convenience we introduce a new coordinate system $z = -Z$, i.e. the z-axis is perpendicular to the OS interface, positive z-values correspond with the semiconductor side. For the electrons in the inversion layer the bent conduction band of the semiconductor together with the step at the interface builds the quantum well. We use the Fang-Howard envelope wave function according to AFS[34]

$$\varphi(z, b) = \begin{cases} \sqrt{\dfrac{b^3}{2}}\, z \exp\left(-\dfrac{bz}{2}\right) & \text{for } z \geq 0 \\ 0 & \text{for } z < 0. \end{cases} \quad (48)$$

The parameter b is varied in order to make the total energy per electron minimal. For the potential several approximations are taken, see App. B for details.

In Fig. 7 the ground state energy ε_0 versus the electron density n_s is shown, ε_0 decreases with decreasing n_s. Now we hold the difference between trap energy and con-
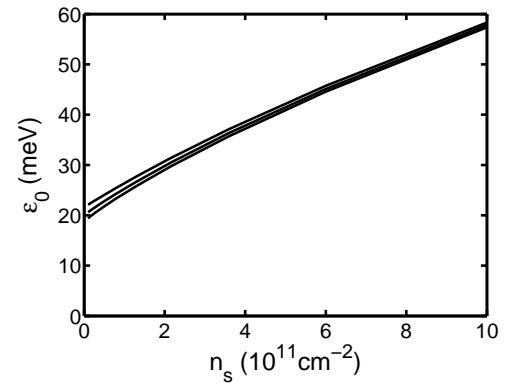

Figure 7. Ground state energy of the inversion layer ε_0 versus the density of electrons in the inversion layer n_s for three temperatures, top-down: $T = 0, 200, 300\,K$.

duction band edge at the interface $\varepsilon_{TsCs} = E_{Ts} - E_{Cs} = E_{Ts} - \mu + \mu_{E_0} + \varepsilon_0$ constant (instead of ε_{Ts0} as before).

When n_s is decreased not only μ_{E_0} but also ε_0 decreases, thus μ drops off faster against E_{Ts} and the transition is more abrupt. This can be seen in Fig. 8a) where results for $\varepsilon_{TsCs} = \text{const.}$ (full lines) and $\varepsilon_{Ts0} = \text{const.}$ (dashed lines) are compared. In both cases Ω_j was calculated numerically. The critical curves $n_s = n_{sc}$ coincide at $T \to 0$ because the values $\varepsilon_{Ts0} \simeq 42.0\,\text{meV}$ and $\varepsilon_{TsCs} \simeq 68.4\,\text{meV}$ were chosen in order to get the same n_{sc} and for higher temperatures because $\varepsilon_0(T)$ is nearly constant over a wide temperature range (Fig. 8c)). For the 3D trap density we assumed again $N_T = 10^{18}\,\text{cm}^{-3}$.

It has to be emphasized that we still use Eq. (1). But the 2D density of positive charges in the depletion layer n_{depl} is a byproduct of the calculation of ε_0. So it is no longer a problem to use Eq. (2) $V_{\text{ins}} \propto n_s + n_{\text{depl}}$ instead of (1) $V_{\text{ins}} \propto n_s$. We will do so henceforward. As a result the slope of the line $E_{Ts} - eV_{\text{ins}}Z/D$ in Fig. 2 is increased, the maximum of $E_T(Z)$ falls off against μ, the $p_+(Z)$-peak gets smaller and narrower. But it is still possible to let coincide the critical curves $\rho(n_s = n_{sc})$ by increasing the trap density N_T in order to compensate the narrower $p_+(Z)$-peak and by increasing ε_{TsCs} in order to get the same critical value n_{sc}. Here it is also important that the 2D charge carrier density of the depletion layer $n_{\text{depl}}(T)$ has almost no n_s dependence for low temperatures as can be seen in Fig. 8d). The resistivity $\rho(T, n_s)$, with $N_T = 2.98 \times 10^{18}\,\text{cm}^{-3}$ and $\varepsilon_{TsCs} \simeq 95.7\,\text{meV}$, for which we get the same n_{sc} as before, is shown in Fig. 8b). The above used depletion density n_{depl} was calculated under the assumption of a background doping density of $N_A = 2 \times 10^{15}\,\text{cm}^{-3}$, which is a typical value for high mobility Si-MOS samples.[35] Here the values for N_T and ε_{TsCs} were chosen in order to get the requested n_{sc} for the given N_A. In reality the value ε_{TsCs} is determined by the chemical nature of the defect and thus the critical density may be different from sample to sample if the background doping is different.

If now the slope of the energy $E_{Ts} - eV_{\text{ins}}Z/D$ is higher due to the inclusion of n_{depl}, the variable $\max(E_T(Z) - \mu)$ which is crucial for the resistivity is less sensitive on V_{ins} and n_s, therefore the transition is less abrupt. The n_s dependence of n_{depl} does not play a role because it hardly exists.

VII. SPATIAL TRAP PROFILE

At higher temperatures the large value of $k_B T$ leads to charged trap states in regions where the chemical potential is even somewhat below the chemical potential (i. e. the $p_+(Z)$-peak broadens) and therefore the resistivity is increased to unrealistic high values, see curves at higher temperature in Fig. 8. But an appreciable density of traps should exist only within the strained region of the oxide,[27] and thus the broadening of the peak beyond the width of this region leads to an unrealistic description. We can resolve this problem by introducing a spatial trap density profile $N_T(Z)$.

If now the trap density N_T is a function of Z it has to remain inside the integral Ω_j (compare Eq. 18)

$$\Omega_j \equiv \int_0^D N_T^+(Z)\, Z^j dZ = \int_0^D N_T(Z)\, p_+(Z)\, Z^j dZ. \tag{49}$$

For simplicity we use here an rectangular spatial trap profile from the OS interface to an arbitrary depth $Z_{\max}$,

$$N_T(Z) = \begin{cases} \frac{n_T}{Z_{\max}} & \text{for } 0 \leq Z \leq Z_{\max} \\ 0 & \text{for } Z > Z_{\max}, \end{cases} \tag{50}$$

where n_T is the 2D trap density. Fig. 9 shows $\rho(T, n_s)$ for $Z_{\max} = 4\,\text{nm}$, the conduction band edge E_{Cs} was used as reference energy, $\varepsilon_{TsCs} \simeq 95.7\,\text{meV}$ was chosen in order to get $n_{sc} = 10^{11}\,\text{cm}^{-2}$. Where the 3D trap density N_T does not vanish its value is assumed to be $2.98 \times 10^{18}\,\text{cm}^{-3}$ as before (see Fig. 8b)), resulting in $n_T = N_T \cdot Z_{\max} = 1.19 \times 10^{12}\,\text{cm}^{-3}$ of which again only a part is charged.

As can be seen in Fig. 9, the behavior for low temperatures has hardly changed, but for high temperatures $\rho(T, n_s)$ now saturates as a broadening of the $p_+(Z)$-peak beyond $Z_{\max}$ does not lead to a further increase in the number of charged scattering centers. This saturation of $\rho(T, n_s)$ is in fairly good agreement with experiments, where ρ for high T is limited as well.

VIII. BROADENING OF THE TRAP ENERGY

As the trap states will not all be identical and in addition the stochastic position distribution will influence them mutually, their energetic position has to be broadened.

We describe the broadening ΔE_T with the help of a normalized distribution function $g\left(\tilde{E}_{Ts}, E_{Ts}, \Delta E_T\right)$ for the trap energy E_{Ts} which characterizes the trap, see Fig. 2. Now E_{Ts} has the meaning of a mean value. (Mean value should not be understood in a strict mathematical sense, e. g. for the Lorentz distribution the mean value does not exist, but in this case it is obvious to take the energy E_{Ts} where the distribution reaches its maximum.) Furthermore $\tilde{E}_{Ts}$ is the value for a particular trap. The probability of $\tilde{E}_{Ts}$ to lie within the interval $[E, E + dE]$ is given by $g(E, E_{Ts}, \Delta E_T)\, dE$. Therefore we replace the probability p_+ of a trap to be charged by

$$P_+(Z) = \int_{-\infty}^{\infty} \frac{g\left(\tilde{E}_{Ts}, E_{Ts}, \Delta E_T\right)}{\frac{1}{2}\exp\left(-\frac{\tilde{E}_{Ts} - eV_{\text{ins}}\frac{Z}{D} - \varepsilon_D \frac{D}{Z} - \mu}{k_B T}\right) + 1}\, d\tilde{E}_{Ts}. \tag{51}$$

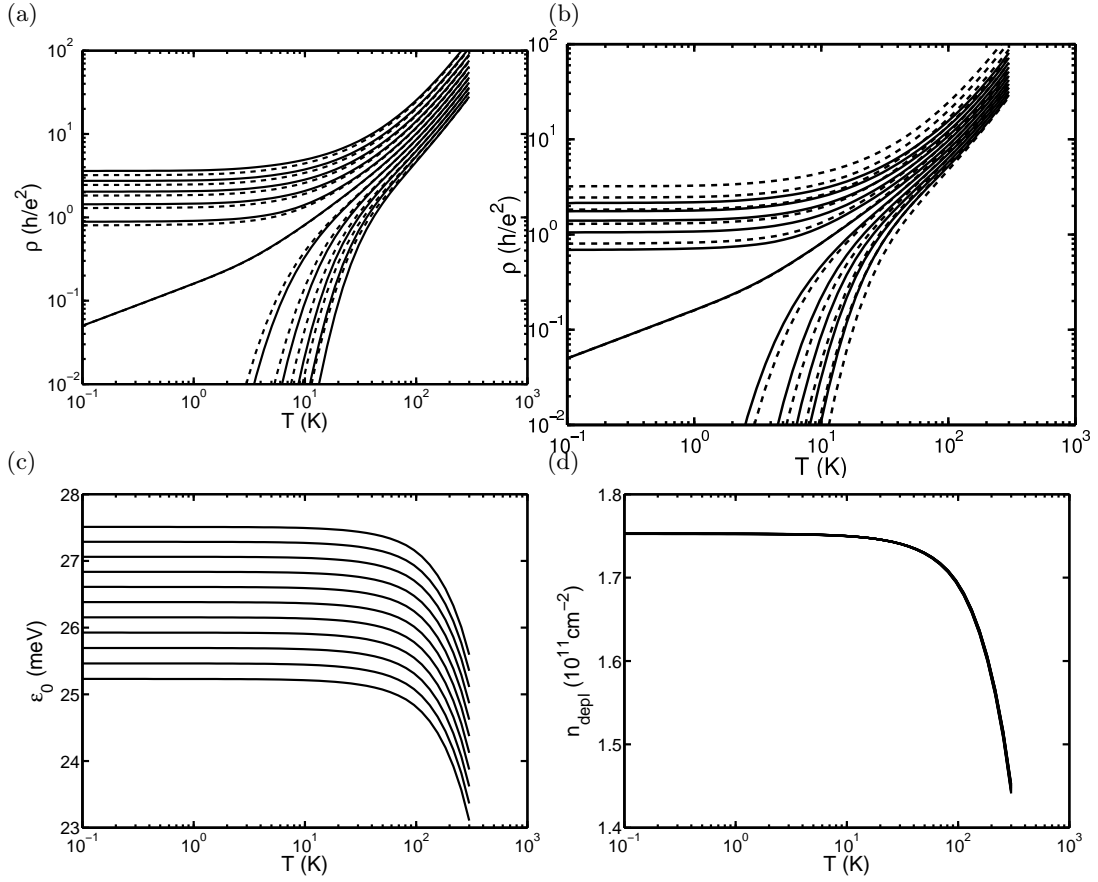The denominator is that of p_+, only E_{Ts} is replaced by

Figure 8. Resistivity $\rho(T, n_s)$ for the potential across the oxide layer a) $V_{\mathrm{ins}} \propto n_s$ with $N_T = 10^{18}\,\mathrm{cm}^{-3}$ and $\varepsilon_{TsCs} \simeq 68.4\,\mathrm{meV}$ and b) $V_{\mathrm{ins}} \propto n_s + n_{\mathrm{depl}}$ with $N_T = 2.98 \times 10^{18}\,\mathrm{cm}^{-3}$ and $\varepsilon_{TsCs} \simeq 95.7\,\mathrm{meV}$. The parameters are chosen to give a critical density of $n_{sc} = 10^{11}\,\mathrm{cm}^{-2}$ for both cases. At the individual curves $n_s = 0.75, \ldots, 1.25 \times n_{sc}$ in steps of $0.05 \times n_{sc}$ from top to bottom, full lines represent $\varepsilon_{TsCs} = \mathrm{const.}$ (realistic case), dashed lines $\varepsilon_{Ts0} = \mathrm{const.}$ (for comparison). In c) the ground state energy of the inversion layer $\varepsilon_0(T, n_s)$ is shown for same n_s values as before, but now assigned bottom up. In d) the depletion density $n_{\mathrm{depl}}(T, n_s)$ is shown, but the n_s dependence vanishes within the line width.
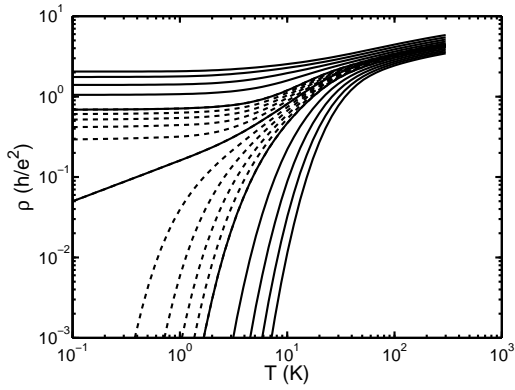

Figure 9. Resistivity $\rho(T, n_s)$ for rectangular spatial trap profile with $Z_{\mathrm{max}} = 4\,\mathrm{nm}$ with curves of constant n_s. The critical density is $n_{sc} = 10^{11}\,\mathrm{cm}^{-2}$ which means that ε_{TsCs} has to be $95.7\,\mathrm{meV}$. Densities for full lines are $n_s = 0.75, \ldots, 1.25 \times n_{sc}$ in steps of $0.05 \times n_{sc}$, and for dashed lines $n_s = 0.96, \ldots, 1.04 \times n_{sc}$ in steps of $0.01 \times n_{sc}$ (always top down), with $N_T = n_T/Z_{\mathrm{max}} = 2.98 \times 10^{18}\,\mathrm{cm}^{-3}$.

$\tilde{E}_{Ts}$. By introducing the dimensionless parameters

$$\alpha = \frac{\Delta E_T}{k_B T}, \tag{52}$$

$$\beta = \frac{E_{Ts} - eV_{\mathrm{ins}}\frac{Z}{D} - \varepsilon_D \frac{D}{Z} - \mu}{k_B T} \tag{53}$$

and a dimensionless distribution function $h(\eta)$ defined by

$$\eta = \frac{\tilde{E}_{Ts} - E_{Ts}}{\Delta E_T}, \tag{54}$$

$$g\left(\tilde{E}_{Ts}, E_{Ts}, \Delta E_T\right) = \frac{1}{\Delta E_T} h\left(\frac{\tilde{E}_{Ts} - E_{Ts}}{\Delta E_T}\right), \tag{55}$$

$$g\left(\tilde{E}_{Ts}, E_{Ts}, \Delta E_T\right) = \frac{1}{\Delta E_T} h(\eta), \tag{56}$$

the probability P_+ can be written as

$$P_+(\alpha, \beta) = \int_{-\infty}^{\infty} \frac{h(\eta)}{\frac{1}{2}\exp(-\alpha\eta - \beta) + 1} d\eta. \tag{57}$$

As a rule this integral cannot be calculated analytically. An exception from this rule is the uniform distribution. If we define the width of the 'rectangle' as $2\Delta E_T$ we get

$$h(\eta) = \begin{cases} \frac{1}{2} & \text{for } -1 < \eta < 1 \\ 0 & \text{elsewhere} \end{cases} \quad (58)$$

and

$$P_+(\alpha,\beta) = \frac{1}{2\alpha} \ln \frac{1 + 2\exp(\beta+\alpha)}{1 + 2\exp(\beta-\alpha)}. \quad (59)$$

We also use the normal distribution

$$h(\eta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\eta^2}{2}\right) \quad (60)$$

with the standard deviation as ΔE_T and the Lorentz distribution (natural line broadening)

$$h(\eta) = \frac{1}{\pi} \frac{1}{\eta^2 + 1} \quad (61)$$

with the half full width at half maximum (FWHM) as ΔE_T.

Fig. 10 shows the numerical results for the three different broadening distributions and different width $\Delta E_T = 0.02$, 0.1 and 0.5 meV. Again we took the conduction band edge at the interface E_{Cs} as reference energy, assumed that traps exist in the oxide only within 4 nm from the OS interface (constant trap density within this region), and used Eq. (2) instead of (1). We chose $\varepsilon_{TsCs} \simeq 95.7$ meV in order to get $n_{sc} = 10^{11}$ cm^{-2} and for the 3D trap density $N_T = n_T/Z_{\max} = 10^{18}$ cm^{-3}.

As one would expect, the transition is less abrupt for higher ΔE_T and does not vanish even for $\Delta E_T = 0.5$ meV. If the broadening is indeed caused by potential fluctuations due to disorder or trap-trap interaction, the value of ΔE_T could even be much larger than $k_B T$ and the transition might be smeared out even stronger.

In the metallic regime the mean trap energy is below the chemical potential. As the normal and the Lorentz distribution have tails, there always remain some charged traps from the upper tail when otherwise all traps would be filled with electrons and therefore would be neutral. On a logarithmic resistivity scale the $\rho(T)$-behavior is changed drastically by the few additional charged traps.

In the insulating regime the mean trap energy is above the chemical potential. Here a large part of the traps is charged and by contrast the few uncharged traps due to the lower tail of the energy distribution hardly play any role.

By means of analytical considerations we got the following estimations for the temperature T_b below which the resistivity becomes almost constant. In general the resistivity ρ varies by some orders of magnitude, so as a criterion for being almost constant we took that region where ρ changes finally by only a factor two down to zero temperature. The markers in Fig. 10 represent these temperatures T_b.

According to this definition, for the insulating behavior $n_s < n_{sc}$ for all three distributions we get

$$T_b = \frac{\max E_T(Z) - E_F}{(4 - \ln 2)k_B} \quad (62)$$

and for the metallic behavior $n_s > n_{sc}$

$$T_b = \begin{cases} \frac{\Delta E_T^2}{(4+\ln 2)k_B(E_F - \max E_T(Z))} & \dots \text{normal distr.} \\ \frac{E_F - \max E_T(Z)}{2(4+\ln 2)k_B} & \dots \text{Lorentz distr.} \end{cases} \quad (63)$$

The uniform distribution has no tails so in the metallic regime there is no temperature range where ρ is almost constant.

IX. CONCLUSIONS

In this work we have performed numerical calculations within the dipole trap model for Si-MOS structures. Originally this model was proposed by Altshuler and Maslov with several approximations, in order to be able to get analytical solutions. Due to our numerical treatment we could eliminate several approximations. We describe the potential inside the insulator by its detailed spatial dependence instead of the saddle-point approximation with the quadratic dependence around its maximum, we fix the trap state energy relative to the conduction band edge instead of relative to the electronic ground state inside the triangular potential well and we have taken into account the detailed change of the chemical potential in the two-dimensional electron layer with respect to the bulk material, which seems to be more realistic than the two cases in the original treatment.

According to our calculations, the metallic regime at high electron densities n_s, where the resistivity is decreasing towards lower temperature, is strongly developed . Also a critical density n_{sc} can be identified with a characteristic temperature dependence different from the metallic and the 'insulating' region. For electron densities $n_s < n_{sc}$, the resistivity curves satturate towards lower temperature and remain constant when the temperature approaches zero. They do not show an insulting behavior in the sense that ρ increases towards zero temperature. Such an increase can in principle be caused e. g. by a further decrease of the chemical potential μ with temperature, as in the work of Altshuler and Maslov for case A, where it was assumed that the temperature dependence of μ ist the same for the two-dimensional electron layer as it is in the Si-bulk material. In the dipol trap model a constant efficient screening is assumed in order to clarify the effects which are caused when the traps change their charge state. A realistic treatment of the temperatue dependence of the electronic screening could also cause an increase in ρ towards lower temperature in that it favors the formation of dipol trap states.
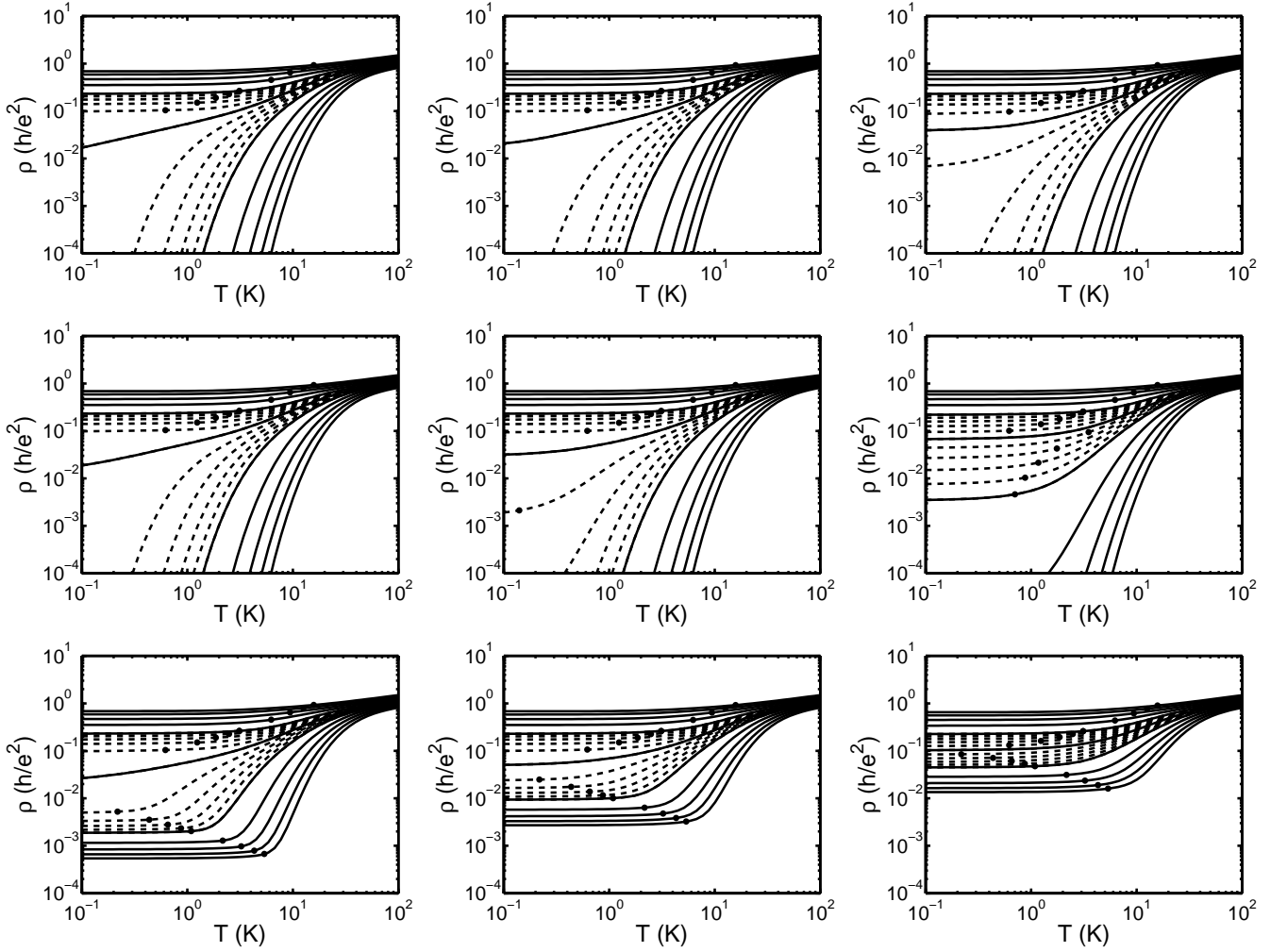
Figure 10. Resistivity $\rho(T, n_s)$ for different types of trap energy broadening functions and different width ΔE_T. The picture columns show uniform distribution, normal distribution, Lorentz distribution from top to bottom. The rows show energy width of $\Delta E_T = 0.02, 0.1, 0.5\,\mathrm{meV}$ from left to right. For the curves n_s is constant. The critical density is $n_{sc} = 10^{11}\,\mathrm{cm}^{-2}$ for $\varepsilon_{TsCs} \simeq 95.7\,\mathrm{meV}$. Full lines show $n_s = 0.75, \ldots, 1.25 \times n_{sc}$ in steps of $0.05 \times n_{sc}$, dashed lines: $n_s = 0.96, \ldots, 1.04 \times n_{sc}$ in steps of $0.01 \times n_{sc}$ (from top to bottom), for a 3D trap density of $N_T = n_T/Z_{\max} = 10^{18}\,\mathrm{cm}^{-3}$.

Furthermore, the quantum corrections in the weak and strong localization regime are also neglected here, but would finally increase the resistance ρ at low T.

In addition, we have further generalized the dipole trap model by dropping the assumptions that the trap states are homogeneously distributed inside the oxide layer and that the energy distribution is δ-like. A narrow spatial distribution of the trap states near the oxide-semiconductor interface limits the number of charged states at high temperatures and thus gives an upper limit for the increase of the resistivity ρ as well. This leads to a good agreement with experimental observations at higher temperatures. The energetic broadening of the trap states on the other hand leads to a finite amount of unoccupied and thus charged states in cases where otherwise all states would lie below the chemical potential μ and the number of charged trap states would go

to zero for $k_B T \to 0$. Thus for high electron densities with metallic behavior the resistivity will not further decrease towards lower temperature, but saturate at a finite values, as has been observed in experiments on Si-MOS structures as well.

The effect of a magnetic field can be taken into account by the Zeeman splitting of the trap states with spin $\pm 1/2$. As shown by Althuler and Maslov, the energetic splitting can turn a metallic behavior into an insulating one. We did not include magnetic field effect in our calculations, but an according energetic shift of the trap states has to lead to the same effects in our refined model as well.

We also like to mention that for low electron densities care has to be taken for the dipol scattering model. It is assumed that the electrons in the two-dimensional layer shield the potential of the charged trap states and thus form together a dipol field which is responsible for the

scattering. At very low electron densities this screening becomes weaker and the scattering will finally increase so that the resistivity should be higher in this regime. These effects have not been taken into account in the frame of the current work, as we like to present the basic effects due to charging of trap states.

Althogether, our detailed numerical calculations within the dipol trap model show that a pronounced metallic state can be caused by trap states at an appropriate energy level inside the oxide of Si-MOS structures. For the realistic assumptions of energetic broadening and narrow spatial distribution near the oxide-semiconductor interface, the behavior is in close agreement with experimental observations.

Appendix A: Temperature behavior of the chemical potential and of the effective electron energy for low temperatures

First we show that

$$\frac{d^n \mu_{E0}(T)}{dT^n}\bigg|_{\varepsilon_{F0}=\text{const.},\,T\to 0} = 0 \quad \text{for } n \geq 1 \qquad \text{(A1)}$$

holds for $\mu_{E0}(T)$ from Eq. (31). In our calculations we use the temperature T and the electron density in the inversion layer n_s as independent variables. So it is allowed to set $n_s = \text{const.}$ and therefore also $\varepsilon_{F0} = \text{const.}$ while varying T, see Eq. (32).

For simplicity we introduce the auxiliary variable

$$x = \frac{k_B T}{\varepsilon_{F0}} \qquad \text{(A2)}$$

and the function

$$M(x) = \frac{\mu_{E0}(T(x)) - \varepsilon_{F0}}{\varepsilon_{F0}}. \qquad \text{(A3)}$$

With the chain rule we find

$$\frac{d^n M(x)}{dx^n} = \frac{\varepsilon_{F0}^{n-1}}{k_B^n} \frac{d^n \mu_{E0}}{dT^n} \quad \text{for } n \geq 1. \qquad \text{(A4)}$$

So Eq. (A1) is equivalent to

$$\frac{d^n M(x)}{dx^n}\bigg|_{x\to 0} = 0 \quad \text{for } n \geq 1. \qquad \text{(A5)}$$

From Eq. (31) we get

$$M(x) = x \ln\left[1 - \exp\left(-\frac{1}{x}\right)\right]. \qquad \text{(A6)}$$

The first two derivatives are

$$\frac{dM}{dx} = \ln\left[1 - \exp\left(-\frac{1}{x}\right)\right] - \frac{1}{\left[\exp\left(\frac{1}{x}\right) - 1\right] x}, \qquad \text{(A7)}$$

$$\frac{d^2 M}{dx^2} = -\frac{1}{\left[\exp\left(\frac{1}{x}\right) - 1\right] x^3} - \frac{1}{\left[\exp\left(\frac{1}{x}\right) - 1\right]^2 x^3}. \qquad \text{(A8)}$$

The second derivative contains only terms of the form

$$g_{a,b} = \frac{1}{\left[\exp\left(\frac{1}{x}\right) - 1\right]^a x^b}. \qquad \text{(A9)}$$

Differentiating $g_{a,b}$ yields

$$\frac{dg_{a,b}}{dx} = a g_{a,b+2} + a g_{a+1,b+2} - b g_{a,b+1}, \qquad \text{(A10)}$$

therefore all higher derivatives $\frac{d^n M}{dx^n}$ also contain only terms $g_{j,k}$ with $j \in \{1, 2, \ldots, n\}$ and $k \in \{n+1, n+2, \ldots, 2n-1\}$. Now we can multiply the numerator and the denominator of $g_{j,k}$ with $\exp(-j/x)$ and write

$$\lim_{x\to 0} g_{j,k}(x) = \lim_{x\to 0} \frac{\exp\left(-\frac{j}{x}\right)}{\left[1 - \exp\left(-\frac{1}{x}\right)\right]^j x^k}$$
$$= \lim_{x\to 0} \frac{\exp\left(-\frac{j}{x}\right)}{x^k}$$
$$= \lim_{y\to\infty} \frac{y^k}{j^k \exp(y)},$$

with $y = j/x$. Applying the rule of L'Hospital k times we get

$$\lim_{x\to 0} g_{j,k}(x) = \lim_{y\to\infty} \frac{k!}{j^k \exp(y)} = 0. \qquad \text{(A11)}$$

The limit for the first derivative also vanishes

$$\lim_{x\to 0} \frac{dM}{dx} = \ln 1 - \lim_{x\to 0} g_{1,1} = 0, \qquad \text{(A12)}$$

so Eq. (A5) holds and

$$\mu_{E_0} \simeq E_{F0} \quad \text{for } k_B T \ll \varepsilon_{F0}. \qquad \text{(A13)}$$

With help of Eq. (A1) we show now that

$$\frac{d^n \bar{\varepsilon}(T)}{dT^n}\bigg|_{\varepsilon_{F0}=\text{const.},\,T\to 0} = 0 \quad \text{for } n \geq 1 \qquad \text{(A14)}$$

holds for Eq. (38).

For not too small η the Fermi-Dirac integral $\mathcal{F}_j(\eta)$ can be approximated by[32]

$$\mathcal{F}_j(\eta) \simeq \frac{\eta^{j+1}}{\Gamma(j+2)}. \qquad \text{(A15)}$$

For low temperatures μ_{E_0} approaches ε_{F0} and therefore it is positive, so from Eq. (38) we get

$$\bar{\varepsilon} \simeq \varepsilon_{F0} \left(\frac{\varepsilon_{F0}}{\mu_{E_0}}\right)^5 \quad \text{for } k_B T \ll \varepsilon_{F0}, \qquad \text{(A16)}$$

and with the chain rule respectively Faá di Bruno's formula

$$\frac{d\bar{\varepsilon}}{dT} = \frac{d\bar{\varepsilon}}{d\mu_{E_0}}\frac{d\mu_{E_0}}{dT}, \tag{A17}$$

$$\frac{d^n\bar{\varepsilon}}{dT^n} = \sum \frac{n!}{k_1!\ldots k_n!}\frac{d^k\bar{\varepsilon}}{d\mu_{E_0}^k}$$
$$\times \left(\frac{1}{1!}\frac{d\mu_{E_0}}{dT}\right)^{k_1}\ldots\left(\frac{1}{n!}\frac{d^n\mu_{E_0}}{dT^n}\right)^{k_n}, \tag{A18}$$
$$k = k_1 + \cdots + k_n,$$

where the sum runs over all integer numbers $k_1,\ldots,k_n \geq 0$ which fulfill

$$k_1 + 2k_2 + \cdots + nk_n = n. \tag{A19}$$

We do not have to find this numbers, we only need to know that for $n > 1$ at least one $k_j > 0$, therefore each term in (A18) contains a factor $\frac{d^j\mu_{E_0}}{dT^j}$, so for $n_s = \text{const.}$, $T \to 0$ the sum vanishes and from Eq. (A16) we get

$$\bar{\varepsilon} \simeq \varepsilon_{F0} \quad \text{for } k_B T \ll \varepsilon_{F0}. \tag{A20}$$

Appendix B: Ground state energy of the inversion layer

As mentioned before we calculate the ground state energy of the inversion layer with help of the Ritz variational principle using the Fang-Howard test envelope wave function[34]

$$\varphi(z,b) = \begin{cases} \sqrt{\frac{b^3}{2}}z\exp\left(-\frac{bz}{2}\right) & \text{for } z \geq 0 \\ 0 & \text{for } z < 0 \end{cases} \tag{B1}$$

and as an approximation for the potential

$$U(z) = \begin{cases} U_d(z) + U_s(z) + U_i(z) & \text{for } z \geq 0 \\ \infty & \text{for } z < 0, \end{cases} \tag{B2}$$

$$U_d(z) \simeq \begin{cases} \frac{e^2 n_{\text{depl}}}{\epsilon_{\text{sc}}\epsilon_0}z\left(1-\frac{z}{2z_d}\right) & \text{for } z < z_d \\ \frac{e^2 n_{\text{depl}}}{2\epsilon_{\text{sc}}\epsilon_0}z_d & \text{for } z > z_d, \end{cases} \tag{B3}$$

$$U_s(z) \simeq \frac{e^2 n_s}{2b\epsilon_{\text{sc}}\epsilon_0}\left[6 - \left((bz)^2 + 4bz + 6\right)\exp\left(-bz\right)\right], \tag{B4}$$

$$U_i(z) \simeq \frac{be^2}{32\pi\epsilon_{\text{sc}}\epsilon_0}\frac{\epsilon_{\text{sc}} - \epsilon_{\text{ins}}}{\epsilon_{\text{sc}} + \epsilon_{\text{ins}}}\frac{1}{4z}. \tag{B5}$$

The term U_d comes from the charged acceptors within the depletion layer with thickness z_d, U_s describes the interaction with all other electrons in the inversion layer, and U_i the interaction with image charges. To write $U_s(z)$ in this form we have to assume that only the first subband is occupied, this is the so called quantum limit. The conduction band edge is built by $U_d(z) + U_s(z)$, therefore the zero point of the energy scale was chosen

to get $U_d(0) + U_s(0) = 0$ which means that the resulting ground state energy is measured against the conduction band edge at the interface E_{Cs} as requested.

The Hamiltonian is given by $\hat{H} = \hat{T} + U$ with the operator for the kinetic energy $\hat{T} = -\frac{\hbar^2}{2m_z}\frac{\partial^2}{\partial z^2}$, where m_z is the z-component of the effective mass of the semiconductor in the bulk. The ground state energy is the expectation value of the Hamiltonian,

$$\varepsilon_0 = \left\langle\hat{T}\right\rangle + \left\langle U_d\right\rangle + \left\langle U_s\right\rangle + \left\langle U_i\right\rangle, \tag{B6}$$

calculated with the value of b which makes the total energy per electron

$$\tilde{\varepsilon} = \left\langle\hat{T}\right\rangle + \left\langle U_d\right\rangle + \frac{1}{2}\left\langle U_s\right\rangle + \left\langle U_i\right\rangle \tag{B7}$$

minimal. The factor $1/2$ in the third term prevents from double counting the electron-electron interaction. With the density $n^* = n_{\text{depl}} + \frac{11}{32}n_s$ introduced by AFS[28] one gets

$$\tilde{\varepsilon} = \frac{\alpha}{2}b^2 + \beta b + \gamma b^{-1} - \frac{\delta(b)}{2}b^{-2}, \tag{B8}$$

$$\alpha = \frac{\hbar^2}{4m_z}, \tag{B9}$$

$$\beta = \frac{e^2}{32\pi\epsilon_{\text{sc}}\epsilon_0}\frac{\epsilon_{\text{sc}} - \epsilon_{\text{ins}}}{\epsilon_{\text{sc}} + \epsilon_{\text{ins}}}, \tag{B10}$$

$$\gamma = \frac{3e^2 n^*}{\epsilon_{\text{sc}}\epsilon_0}, \tag{B11}$$

$$\delta = \frac{12e^2(N_A - N_D)}{\epsilon_{\text{sc}}\epsilon_0} \tag{B12}$$

$$\times \left\{1 - \left[\frac{(bz_d)^2}{12} + \frac{bz_d}{2} + 1\right]\exp\left(-bz_d\right)\right\}. \tag{B13}$$

N_A and N_D are the densities of the acceptors and donors respectively. The coefficients have been chosen in order to get α, β, γ, $\delta > 0$ and to get a most simple equation

$$\frac{d\tilde{\varepsilon}}{db} = \alpha b + \beta - \gamma b^{-2} + \delta(b)b^{-3} = 0. \tag{B14}$$

As AFS used $\frac{e^2 n_{\text{depl}}}{\epsilon_{\text{sc}}\epsilon_0}z\left(1-\frac{z}{2z_d}\right)$ for $0 < z < \infty$ instead of Eq. (B3) they did not get the term proportional to $\exp\left(-bz_d\right)$, under normal circumstances it is very small but formally it is necessary to see that $b \to 0$, $\tilde{\varepsilon} \to -\infty$ is not the global minimum. Furthermore they neglected β and $\delta(b)b^{-3}$ and got

$$b = \left(\frac{\gamma}{\alpha}\right)^{1/3} = \left(\frac{12e^2 n^* m_z}{\epsilon_{\text{sc}}\epsilon_0\hbar^2}\right)^{1/3}. \tag{B15}$$

We neglected only $\delta(b) b^{-3}$, this leads to

$$b = F\left(K + \frac{1}{K} - 1\right), \tag{B16}$$

$$F = \frac{\beta}{3\alpha} = \frac{e^2 m_z}{24\pi\epsilon_{\rm sc}\epsilon_0\hbar^2}, \tag{B17}$$

$$K = \left[\frac{1}{2}\left(3B - 2 + \sqrt{3B(3B-4)}\right)\right]^{1/3}, \tag{B18}$$

$$B = \frac{9\alpha^2\gamma}{\beta^3} = \frac{\gamma}{\beta F^2} = \frac{96\pi n^*}{F^2}\frac{\epsilon_{\rm sc} + \epsilon_{\rm ins}}{\epsilon_{\rm sc} - \epsilon_{\rm ins}}. \tag{B19}$$

The density $n_{\rm depl}$ can be calculated from the total band bending $e\phi_0 = E_{Cb} - E_{Cs}$ (b=bulk, s=surface),

$$en_{\rm depl} = e(N_A - N_D)z_d = \sqrt{2(N_A - N_D)\epsilon_{\rm sc}\epsilon_0 e\phi_{0{\rm d}}}, \tag{B20}$$

$$e\phi_{0d} = e\phi_0 - e\phi_{0s} - k_B T. \tag{B21}$$

Here $e\phi_{0{\rm d}}$ is the band bending caused by the charges in the depletion layer (acceptors and donors) and $e\phi_{0s}$ that caused by the free electrons in the inversion layer. To get this equations one has to assume that in the depletion layer all acceptors are charged and that in the bulk there is charge neutrality. The boundary between the depletion layer and the bulk is not sharp, this is described by the term $-k_B T$ in $e\phi_{0{\rm d}}$.[36] To calculate ϕ_{0s} one has to solve the Poisson equation with the charge density which corresponds with the test wave function φ, the result is

$$\phi_{0s} = \frac{3en_s}{b\epsilon_{\rm sc}\epsilon_0}. \tag{B22}$$

As can be seen from Fig. 4 for the total band bending

$$e\phi_0 = E_{Cb} - E_{Cs},$$
$$e\phi_0 = E_{Cb} - \mu + \mu - E_0 + E_0 - E_{Cs},$$
$$e\phi_0 = -\mu_{Cb} + \mu_{E_0} + \varepsilon_0 \tag{B23}$$

holds. The chemical potential relative to the conduction band edge in the bulk μ_{Cb} is determined by the charge neutrality and can be calculated by solving

$$N_C\mathcal{F}_{1/2}\left(\frac{\mu_{Cb}}{k_B T}\right) + \frac{N_A}{g_A\exp\left(\frac{\varepsilon_A - \varepsilon_g - \mu_{Cb}}{k_B T}\right) + 1} =$$
$$= N_V\mathcal{F}_{1/2}\left(\frac{-\varepsilon_g - \mu_{Cb}}{k_B T}\right) \tag{B24}$$

numerically. N_C and N_V are the effective densities of state in the conduction band and in the valence band, ε_A is the acceptor ionization energy, g_A the acceptor degeneracy factor, and ε_g the gap energy. According to Sze[27] this quantities are given by

$$N_C = g_s g_{\rm v3D}\left(\frac{m_{\rm de3D}k_B T}{2\pi\hbar^2}\right)^{3/2}, \tag{B25}$$

$$N_V = g_s\left(\frac{m_{\rm dh3D}k_B T}{2\pi\hbar^2}\right)^{3/2}, \tag{B26}$$

$$\varepsilon_g = \varepsilon_{\rm g0Si} - \frac{\alpha_{\rm Si}T^2}{T + \beta_{\rm Si}}. \tag{B27}$$

The values we used can be found in Table I.

The ground state energy ε_0 itself is a (small) part of the the total band bending, this problem is solved by a fix point iteration using $\varepsilon_0 = 0$ as start value.

[*] gerhard.brunthaler@jku.at

[1] S. V. Kravchenko, G. V. Kravchenko, J. E. Furneaux, V. M. Pudalov, and M. DIorio, Phys. Rev. B **50**, 8039 (1994).

[2] S. V. Kravchenko, W. E. Mason, G. E. Bowker, J. E. Furneaux, V. M. Pudalov, and M. DIorio, Phys. Rev. B **51**, 7038 (1995).

[3] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979).

[4] E. Abrahams, S. V. Kravchenko, and M. P. Sarachik, Rev. Mod. Phys. **73**, 251 (2001).

[5] M. Y. Simmons, A. R. Hamilton, M. Pepper, E. H. Linfield, P. D. Rose, D. A. Ritchie, A. K. Savchenko, and T. G. Griffiths, Phys. Rev. Lett. **80**, 1292 (1998).

[6] M. P. Lilly, J. L. Reno, J. A. Simmons, I. B. Spielman, J. P. Eisenstein, L. N. Pfeiffer, K. W. West, E. H. Hwang, and S. Das Sarma, Phys. Rev. Lett. **90**, 056806 (2003).

[7] V. Senz, T. Ihn, T. Heinzel, K. Ensslin, G. Dehlinger, D. Grützmacher, U. Gennser, E. H. Hwang, and S. D. Sarma, Physica E **13**, 723 (2002).

[8] S. J. Papadakis and M. Shayegan, Phys. Rev. B **57**, R15068 (1998).

[9] F. Stern, Phys. Rev. Lett. **44**, 1469 (1980).

[10] A. Gold and V. T. Dolgopolov, Phys. Rev. B **33**, 1076 (1986).

[11] S. Das Sarma, Phys. Rev. B **33**, 5401 (1986).

[12] A. Gold, J. Phys.: Cond. Matt. **15**, 217 (2003).

[13] S. Das Sarma, M. P. Lilly, E. H. Hwang, L. N. Pfeiffer, K. W. West, and J. L. Reno, Phys. Rev. Lett. **94**, 136401 (2005).

[14] S. Das Sarma and E. Hwang, Solid State Comm. **135**, 579 (2005).

[15] A. M. Finkel'stein, Z. Phys. B Condes. Matter **56**, 189 (1984).

[16] C. Castellani, C. Di Castro, P. A. Lee, and M. Ma, Phys. Rev. B **30**, 527 (1984).

[17] A. Punnoose and A. M. Finkel'stein, Phys. Rev. Lett. **88**, 016802 (2001).

[18] A. Punnoose and A. M. Finkel'stein, Science **310**, 289 (2005).

[19] G. Zala, B. N. Narozhny, and I. L. Aleiner, Phys. Rev. B **64**, 214204 (2001).

[20] I. V. Gornyi and A. D. Mirlin, Phys. Rev. B **69**, 045313 (2004).

[21] S. V. Kravchenko and M. P. Sarachik, Rep. Prog. Phys. **67**, 1 (2004).

[22] A. A. Shashkin, Physics-Uspekhi **48**, 129 (2005).

[23] V. T. Dolgopolov, Low Temp. Phys. **33**, 98 (2007).

[24] F. Evers and A. D. Mirlin, Rev. Mod. Phys. **80**, 1355 (2008).

[25] W. R. Clarke, C. E. Yasin, A. R. Hamilton, A. P. Micolich, M. Y. Simmons, K. Muraki, Y. Hirayama, M. Pepper, and D. A. Ritchie, Nat Phys **4**, 55 (2008).

[26] B. L. Altshuler and D. L. Maslov, Phys. Rev. Lett. **82**, 145 (1999).

[27] S. M. Sze, *Physics of semiconductor devices*, 2nd ed. (New York, Wiley-Interscience, 1981).

[28] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).

[29] T. Hori, *Gate Dielectrics and MOS ULSIs* (Springer Verlag, Berlin, 1997).

[30] T. M. Klapwijk and S. Das Sarma, Solid State Comm. **110**, 581 (1999).

[31] Typo in Eq. 8 of AM paper (D. L. Maslov; private communication).

[32] J. S. Blakemore, *Semiconductor Statistics* (New York, Courier Dover Publications, 1987).

[33] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Brooks/Cole, 1976).

[34] F. F. Fang and W. E. Howard, Phys. Rev. Lett. **16**, 797 (1966).

[35] V. M. Pudalov, private communication.

[36] F. Stern, Phys. Rev. B **5**, 4891 (1972).